nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-019-0909-z

Palaeoproteomics resolves sloth relationships

Samantha Presslee[1,2,3,24], Graham J. Slater[4,24], François Pujos[5], Analía M. Forasiepi[5], Roman Fischer[6], Kelly Molloy[7], Meaghan Mackie[3,8], Jesper V. Olsen[8], Alejandro Kramarz[9], Matías Taglioretti[10], Fernando Scaglia[10], Maximiliano Lezcano[11], José Luis Lanata[11], John Southon[12], Robert Feranec[13], Jonathan Bloch[14], Adam Hajduk[15], Fabiana M. Martin[16], Rodolfo Salas Gismondi[17], Marcelo Reguero[18], Christian de Muizon[19], Alex Greenwood[20,21], Brian T. Chait[7], Kirsty Penkman[22], Matthew Collins[3,23] and Ross D. E. MacPhee[2]*

The living tree sloths *Choloepus* and *Bradypus* are the only remaining members of Folivora, a major xenarthran radiation that occupied a wide range of habitats in many parts of the western hemisphere during the Cenozoic, including both continents and the West Indies. Ancient DNA evidence has played only a minor role in folivoran systematics, as most sloths lived in places not conducive to genomic preservation. Here we utilize collagen sequence information, both separately and in combination with published mitochondrial DNA evidence, to assess the relationships of tree sloths and their extinct relatives. Results from phylogenetic analysis of these datasets differ substantially from morphology-based concepts: *Choloepus* groups with Mylodontidae, not Megalonychidae; *Bradypus* and *Megalonyx* pair together as megatherioids, while monophyletic Antillean sloths may be sister to all other folivorans. Divergence estimates are consistent with fossil evidence for mid-Cenozoic presence of sloths in the West Indies and an early Miocene radiation in South America.

The sloths (Xenarthra, Folivora), nowadays a taxonomically narrow (six species in two genera) component of the fauna of South and Central America[1,2], were once a highly successful clade of placental mammals as measured by higher-level diversity (Fig. 1). Diverging some time in the Palaeogene from their closest relatives, the anteaters (Vermilingua), folivorans greatly expanded their diversity and range, eventually reaching North America as well as the West Indies[3–8]. During the late Cenozoic, sloth lineage diversity may have expanded and contracted several times[9]. Final collapse occurred in the late Quaternary (end-Pleistocene on the continents, mid-Holocene in the West Indies), leaving only the lineages that culminated in the extant two-toed (*Choloepus*) and three-toed (*Bradypus*) tree sloths.

Radically differing from other sloth taxa in their manifold adaptations for 'inverted' suspensory locomotion, tree sloths have an obscure evolutionary history[10]. Despite their overall similarity in body plans, tree sloths probably acquired their remarkable locomotor adaptations separately, one of many indications that the course of folivoran evolution has been marked by detailed convergences among evolutionarily distinct clades[11–19]. The current consensus[8–10,16,17] in morphology-based phylogenetic treatments is to place the three-toed sloth as sister to all other folivorans (Fig. 1, Eutardigrada), while *Choloepus* is typically nested within the otherwise extinct family Megalonychidae, either proximate to or actually within the group that radiated in the West Indies[3,7,11,13,16,20,21]. Although this arrangement recognizes the existence of convergence in the origins of arboreality in tree sloths, it has proved difficult to test effectively. Sloth palaeontology is an active field of enquiry (for example, refs. [10,17,21–30]), but the placement of a number of early Neogene clades is uncertain or disputed[31] (signified by 'unallocated basal folivorans' in Fig. 1), and the nature of their relationships with the tree sloths is accordingly indeterminate. This has an obvious impact on our ability to make macro-evolutionary inferences[14] (for example, ancestral modes of locomotion) for tree sloth species, which have no known pre-Quaternary fossil record[10].

Genomic evidence, now routinely used in mammalian systematic research and phylogenetic reconstruction, has so far been of limited use in evaluating these issues. Mitochondrial and at least some nuclear sequence data are available for most well-defined species of living tree sloths, but published ancient DNA (aDNA)

[1]Department of Archaeology and BioArCh, University of York, Heslington, UK. [2]Department of Mammalogy, American Museum of Natural History, New York, NY, USA. [3]Paleoproteomics Group, Natural History Museum of Denmark and University of Copenhagen, Copenhagen, Denmark. [4]Department of the Geophysical Sciences, University of Chicago, Chicago, IL, USA. [5]Instituto Argentino de Nivología, Glaciología y Ciencias Ambientales, CCT-CONICET-Mendoza, Mendoza, Argentina. [6]Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [7]Chait Laboratory and National Resource for the Mass Spectrometric Analysis of Biological Macromolecules, The Rockefeller University, New York, NY, USA. [8]Novo Nordisk Foundation Center for Protein Research, Copenhagen, Denmark. [9]Sección Paleovertebrados, Museo Argentino de Ciencias Naturales 'Bernardino Rivadavia', Buenos Aires, Argentina. [10]Museo Municipal de Ciencias Naturales 'Lorenzo Scaglia', Mar del Plata, Argentina. [11]Instituto de Investigaciones en Diversidad Cultural y Procesos de Cambio, CONICET and Universidad Nacional de Río Negro, Bariloche, Argentina. [12]Keck-CCAMS Group, Earth System Science Department, University of California, Irvine, Irvine, CA, USA. [13]New York State Museum, Albany, NY, USA. [14]Florida Museum of Natural History, University of Florida, Gainesville, FL, USA. [15]Museo de la Patagonia 'F. P. Moreno', Bariloche, Argentina. [16]Centro de Estudios del Hombre Austral, Instituto de la Patagonia, Universidad de Magallanes, Punta Arenas, Chile. [17]BioGeoCiencias Lab, Facultad de Ciencias y Filosofía/CIDIS, Universidad Peruana Cayetano Heredia, Lima, Peru. [18]CONICET and División Paleontología de Vertebrados, Museo de La Plata. Facultad de Ciencias Naturales, Universidad Nacional de La Plata, La Plata, Argentina. [19]Centre de Recherches sur la Paléobiodiversité et les Paléoenvironnements, Muséum national d'Histoire naturelle, Paris, France. [20]Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany. [21]Department of Veterinary Medicine, Freie Universität Berlin, Berlin, Germany. [22]Department of Chemistry, University of York, Heslington, UK. [23]McDonald Institute for Archaeological Research, University of Cambridge, Cambridge, UK. [24]These authors contributed equally: S. Presslee, G. J. Slater. *e-mail: macphee@amnh.org

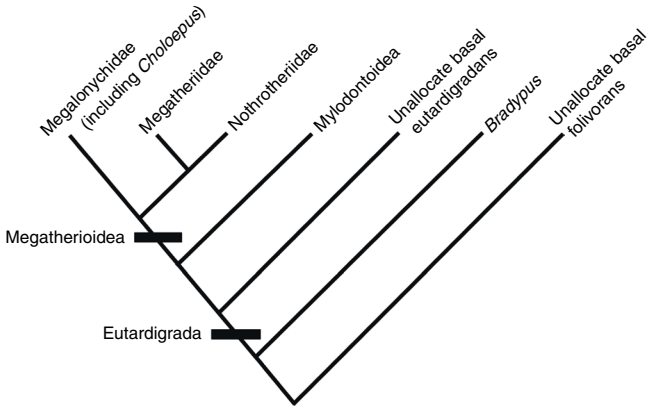

Fig. 1 | Phylogenetic relationships among major folivoran taxa based on morphological evidence, with existence of unallocated taxa acknowledged. In this framework, the three-toed tree sloth *Bradypus* is sister to other sloths (grouped here as Eutardigrada), while the two-toed tree sloth *Choloepus* is included within Megalonychidae. Morphological evidence mostly after refs. [8,16].

evidence exists for only two late Pleistocene species[32–35]. Lack of aDNA evidence is not surprising, given that the vast majority of sloth species lived in temperate or tropical environments not conducive to aDNA preservation. Despite these limitations, aDNA analyses have tentatively pointed to a set of relationships between extant sloths and their extinct relatives that are very different from those implied by morphological data: the three-toed sloth is consistently recovered in association with the North American megatherioid *Nothrotheriops shastensis*[33,36,37], a position reflected in some older classifications[13,20,38], while the two-toed sloth is firmly established as sister to the South American mylodontoid *Mylodon darwinii*[33–37,39,40]. This information, however, is not sufficient for rigorous testing, with molecular evidence, of cladistic relationships established solely on morphological grounds.

There is another potential source of ancient biomolecular evidence: sequence information derived from proteins[41–44]. Because an organism's proteins are coded by its DNA, amino acid sequences in a protein are directly controlled by the gene sequences which specify them. Importantly, proteins—especially structural proteins like collagen and myosin—characteristically degrade at a slower rate than DNA[45–47]. Using tandem mass spectrometry coupled with high-performance liquid chromatography, it has proved possible to recover authentic collagen sequence information from mammalian fossils as old as mid-Pliocene (3.5–3.8 Ma)[48], which exceeds the current aDNA record (560–780 kyr BP) by a substantial interval[49,50]. Another advantage is that proteomic data can potentially be recovered from specimens from a wide range of taphonomic contexts, including those generally inimical to aDNA preservation[51]. There are of course limitations. Bones and teeth are typically the only parts of vertebrate bodies that preserve as fossils, which restricts the choice of proteins to those that occur in large amounts in such tissues. Type 1 collagen comprises ~90% of the organic fraction of vertebrate bone[52] and is the only bone protein[46] that is well represented in taxonomically extensive libraries such as the National Center for Biotechnology Information (NCBI). Since type 1 collagen is coded by only two genes, *COL 1A1* and *COL 1A2*, only a small fraction of a species' genome can be accessed with this probe. In the context of palaeontology, phylogenetic analyses of type 1 collagen have been shown to yield results that are highly congruent with those produced by aDNA, especially at higher taxonomic levels[43,53].

One such application is the testing of morphology-based hypotheses of higher-level relationships where there is a strong possibility that pervasive homoplasy among and between target groups has affected morphological character analysis and therefore classification, as in the case of incorrectly homologized caniniform tooth loci in living tree sloths[54]. Because dental features have always played a large role in folivoran systematics[7,10,12,13,16,30], such fundamental reinterpretations are likely to have a major impact. Clearly, it is desirable to use as many sources of inference as possible in reconstructing phylogeny. Also, molecular data lend themselves well to estimating divergence timing of major clades—another critical problem in folivoran systematics[28,33,34].

Results

To address some of the questions raised in the previous section, as well as to add to the available molecular database for folivorans, we utilized proteomic data collected from fossil and living sloths to focus on three fundamental issues: (1) relationships of tree sloths to each other and to other folivorans; (2) composition of folivoran superfamilies Megatherioidea and Mylodontoidea; and (3) divergence dating of major sloth ingroups. Results were tested against datasets that additionally incorporated published genomic and phenomic information.

Samples. A total of 120 xenarthran samples comprising 24 different genus-level taxa (see Supplementary information, Supplementary Table 1.) were screened for protein survival using both amino acid racemization (AAR) and matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry. Three additional xenarthran sequences were taken from the literature (see Methods, Proteomic analysis). Of these, 34 or 28.3% of the total number of samples (including 31.0% of 103 folivoran samples) produced promising results with both AAR and MALDI-TOF mass spectrometry. From these, the best sample per taxon was selected for liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis to derive protein sequences, with some additions to maximize taxonomic coverage (Fig. 2 and Table 1). We resampled the specimen of *Megatherium* previously utilized in ref. [44]; the results presented here are de novo. The samples of *Neocnus dousman* and *Megalocnus zile* did not pass both MALDI-TOF and AAR screening criteria, but it was decided to analyse them because they were the best representatives of their species. However, because coverage for the *Megalocnus* sample was particularly poor, the recovered sequence being mostly contaminants, it was not used in the phylogenetic analyses. To provide modern comparisons, samples of *Bradypus variegatus* (AMNH 20820) and *Choloepus hoffmanni* (AMNH 139772) were also subjected to LC–MS/MS analysis. For further details on all samples, see Supplementary information, especially Supplementary Table 1. Relevant procedures for recovery of sequence information and estimation of phylogenetic relationships are presented in Methods.

Samples ranged in assigned age from late Miocene to mid-Holocene (Supplementary Table 1), but the 19 samples successfully screened are all Quaternary (Table 1). Of these, 15 were selected for radiocarbon dating and 10 returned finite ^{14}C ages (Supplementary Table 2). The oldest specimen that yielded sequence information, *Glossotherium robustum* MACN-PV 2652, is catalogued as Bonaerian SALMA (South American land mammal age, 128–400 ka[55]), but this age assignment cannot be independently confirmed.

To keep nomenclature manageable, we make frequent reference to the relatively simple traditional taxonomic scheme presented in Fig. 1, which is in turn based on a large simultaneous analysis of folivoran relationships[8,16]. Major departures from traditional frameworks will be denoted where necessary by an asterisk, but only for formal taxonomic names (for example, *Mylodontoidea, that is, clade redefined to include *Choloepus*, not a traditional member).

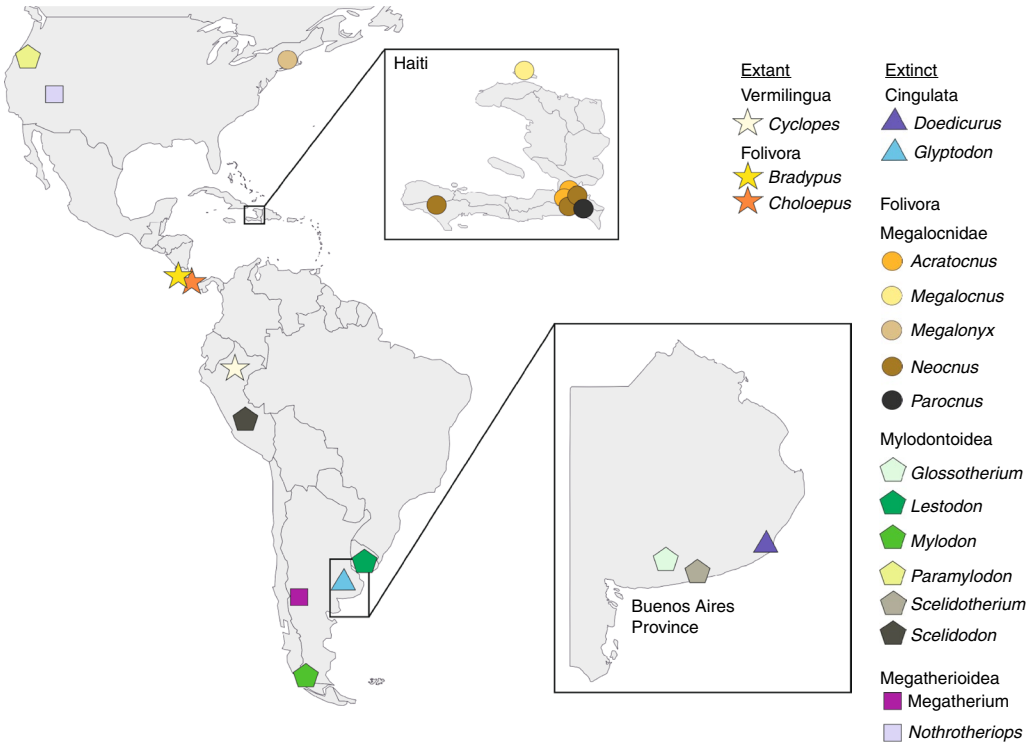

Fig. 2 | Geographical locations of sequenced samples. Sequences for *Cyclopes* and *Lestodon* (in bold) taken from the literature; other taxa, this paper (Table 1, Supplementary Table 1).

Table 1 | Collagen peptides and percentage coverage of the sequenced ancient and modern samples

Museum reference[a]	ID	Species	Collagen peptides (*n*)	Coverage (%)
MMP 5672	15191	*Doedicurus* sp.	867	90
MACN-PV 7	15194	*Glyptodon* sp.	731	84
UF 76796	15559	*Acratocnus ye*	696	86
UF 76385	15565	*Acratocnus ye*	629	87
AMNH-M 20820	16265	*Bradypus variegatus*	793	88
AMNH-M 139772	17009	*Choloepus hoffmanni*	1,109	94
MACN-PV 2652	15216	*Glossotherium robustum*	837	88
UF 169931	15564	*Megalocnus zile*[b]	6	6
NYSM VP-46	16849	*Megalonyx jeffersonii*[c]	874	85
MAPBAR 3965	15225	*Megatherium americanum*	520	81
UMAG ah 5854	16222	*Mylodon darwinii*	1,371	96
UF 171347	15548	*Neocnus comes*	699	84
UF 170210	15780	*Neocnus comes*	591	84
UF 75469	15781	*Neocnus dousman*	614	74
USNM 244372	14723	*Nothrotheriops shastensis*	528	79
USNM 3000	14715	*Paramylodon harlani*	642	87
UF 75526	15556	*Parocnus serus*	575	82
MUSM 1386	17480	*Scelidodon* sp.	1,324	92
MACN-PV 1791	15202	*Scelidotherium* sp.	475	76

[a]Institutional acronyms: AMNH-M, American Museum of Natural History (Mammalogy), New York, NY, USA; MACN-PV, Museo Argentino de Ciencias Naturales 'Bernardino Rivadavia', Buenos Aires, Argentina; MAPBAR, Museo de la Asociación Paleontológica Bariloche, Bariloche, Argentina; MMP, Museo Municipal de Ciencias Naturales 'Lorenzo Scaglia' Mar del Plata, Argentina; MUSM, Museo de Historia Natural, Universidad Nacional Mayor de San Marcos, Lima, Peru; NYSM VP, New York State Museum (Vertebrate Paleontology), Albany, NY, USA; UF, University of Florida, Natural History Museum of Florida, Gainesville, FL, USA; UMAG ah, Instituto de La Patagonia, Universidad de Magallanes, Punta Arenas, Chile; USNM, United States National Museum of Natural History (Paleobiology), Washington DC, USA. [b]Mainly contaminants; not sequenced. [c]SDS–PAGE protein extraction.

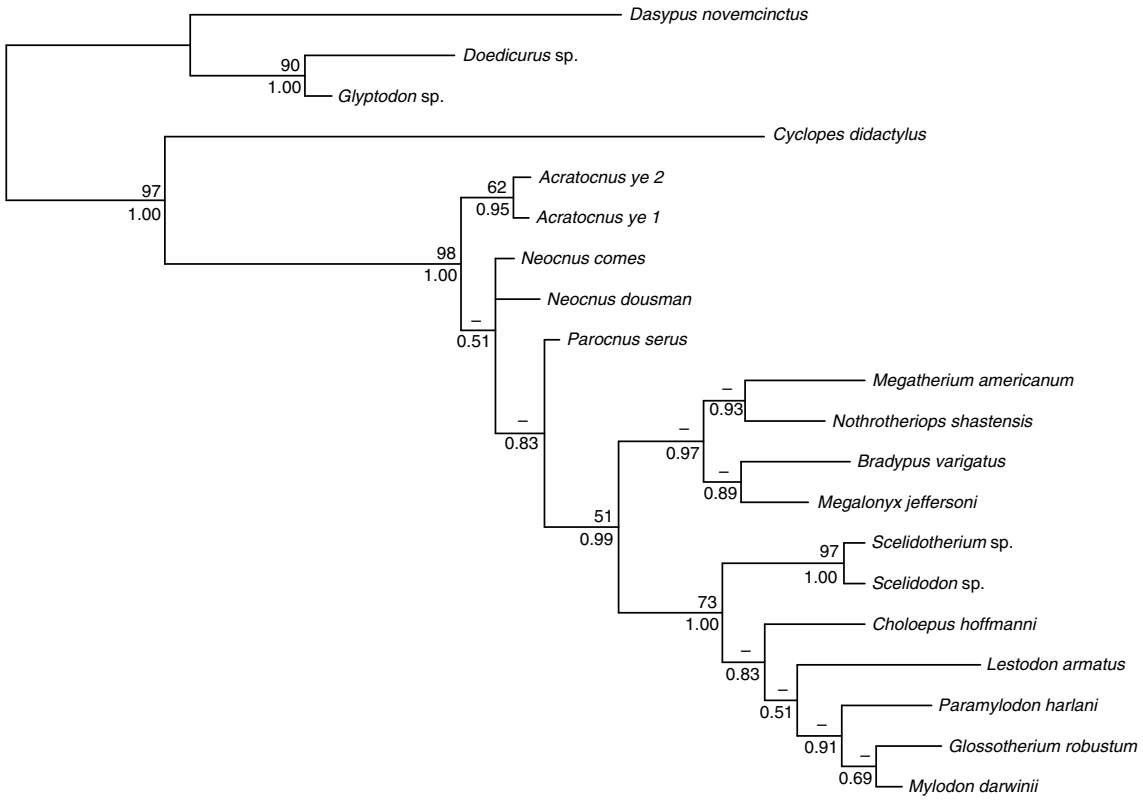

Fig. 3 | Fifty percent majority rule consensus tree from Bayesian analysis of the proteomic data without temporal information, as performed in MrBayes. Values below nodes are posterior probabilities for the descendant clade (see Results, 'Phylogenetic reconstruction'). Values above nodes are bootstrap support derived from 10,000 bootstrap replicates. A dash (–) indicates that a node was not represented in the 50% majority rule bootstrap consensus. Extant *Dasypus* and extinct *Doedicurus* and *Glyptodon* are members of the order Cingulata; extant *Cyclopes* is a representative of Vermilingua, which, together with Folivora, comprise the order Pilosa. Cingulates and pilosans together comprise the superorder Xenarthra (see also Fig. 4).

Phylogenetic reconstruction. Parsimony and Bayesian topology searches resulted in largely congruent topologies. Bootstrap support (BS) under parsimony was generally low, as might be expected given few variable sites, while Bayesian posterior probabilities (PP), which make full use of the data, resulted in somewhat higher clade support (Fig. 3 and Supplementary Fig. 2). Although Antillean sloth relationships are not meaningfully resolved, other folivorans assort into two reciprocally monophyletic clades (PP = 0.99) that are consistent with aDNA results[33,34]. The first includes the three-toed sloth and various extinct taxa traditionally considered megatherioid (PP = 0.97). The sister group relationship of *Megatherium* and *Nothrotheriops* (PP = 0.93) is non-controversial (Fig. 1), but in the Bayesian consensus we unexpectedly recovered a previously unreported and moderately well-supported pairing of *Megalonyx* with *Bradypus* (PP = 0.89) (see Discussion). The second monophyletic clade (BS = 73, PP = 1.00) consists of traditional mylodontoids plus *Choloepus*. Because the inclusion of *Choloepus* in this group contrasts markedly with results achieved using morphological datasets, we designate this clade as *Mylodontoidea. Here, *Scelidotherium* + *Scelidodon* is the earliest diverging branch and *Choloepus* is recovered as part of a clade (PP = 0.83), consistent with accepted mylodontid interrelationships[16,30,56].

To further interrogate the reliability of our proteomic topologies, we concatenated our collagen sequences with previously published mitochondrial genome sequences (hereafter, proteomic + genomic data) for all extant folivorans (two species of *Choloepus*, four species of *Bradypus*), two extinct folivorans (*M. darwinii* and *N. shastensis*) and the two extant outgroup taxa[33,34]. Bayesian analysis (Supplementary Fig. 2) of the combined dataset yielded a topology almost identical to that recovered using proteomic data alone, but in this instance *Megatherioidea (including *Bradypus*) and *Mylodontoidea (including *Choloepus*) were unambiguously recovered as reciprocally monophyletic clades (PP = 1). Recovery of a paraphyletic *Bradypus* (with respect to *Megalonyx*) is almost certainly due to a long genomic branch and lack of proteomic data for *Bradypus torquatus*, combined with a comparable lack of genomic data for *Megalonyx*. As the monophyly of *Bradypus* has never been questioned and this result is based exclusively on relative branch length, we constrained *Bradypus* monophyly for subsequent analyses, though analyses without a constraint were not noticeably different.

Molecular clock considerations and divergence time estimates. Incorporating time as an analytical component in analysis of the combined dataset yielded a well-supported and monophyletic Antillean clade (PP > 0.99), although within-clade relationships were not satisfactorily resolved. More unexpectedly, in light of traditional taxonomic concepts, BEAST placed the Antillean clade as a well-supported sister to *Megatherioidea plus *Mylodontoidea (PP = 0.97) rather than pairing it with one or the other. Support for megatherioid (PP > 0.99) and mylodontoid (PP > 0.99) monophyly remained strong, but variable for constituent sub-clades.

The relatively permissive constraints employed for calculation of divergences make it difficult to draw detailed conclusions regarding the tempo of sloth diversification, although mean ages in the combined analysis are reasonably consistent with inferences based on both genomic[33,34] and morphological[16] data (Fig. 4 and Table 2). Posterior mean node ages suggest an early Oligocene origin for

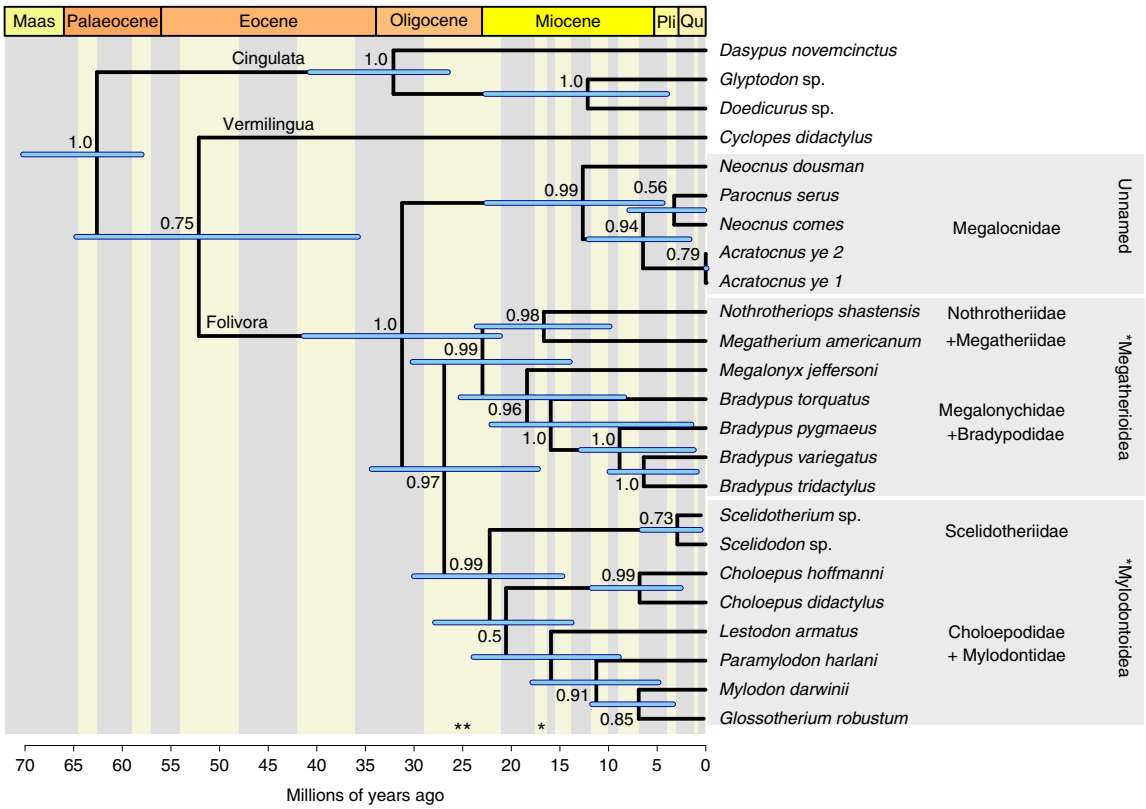

Fig. 4 | Time-scaled maximum clade credibility tree from BEAST analysis of 24 extant and extinct xenarthran collagen sequences plus published mitochondrial genomes. Branch lengths are mean values from the retained posterior sample, while blue bars represent 95% highest posterior density intervals. Values at nodes are posterior probabilities (note that the monophyly of *Bradypus* is constrained here). Vertical shaded bars correspond to SALMAs, two of which are emphasized: Deseadan (**), 29–21 Ma, during which the first generally accepted representatives of traditional Megatherioidea and Mylodontoidea appear palaeontologically; and the Santacrucian (*), 17.5–16.3 Ma, the SALMA during which mylodontids maintained substantial taxonomic diversity but megalonychids and megatheriids declined[9]. On the right (grey boxes), folivoran species used in analyses are associated with their traditional family names, but with superfamily contents organized according to phylogenetic conclusions in the text. Megalocnidae is placed outside traditional superfamily structure in its own (unnamed) box. The tree implies that the fundamental split within Folivora is not between Megatherioidea and Mylodontoidea versus Bradypodoidea as classically understood, but rather between redefined *Megatherioidea and *Mylodontoidea versus Megalocnidae. Maas, Maastrichian; Pli, Pliocene; Qu, Quaternary.

folivorans, with megatherioids and mylodontoids diverging in the middle to late Oligocene (Deseadan SALMA) and the generally recognized families originating within the middle Miocene (Colloncuran-Laventan SALMAs). The combined analysis indicates that the last time *Choloepus* and *Bradypus* shared a common ancestor was ~26.9 Ma (95% highest posterior density interval, 17.2–34.4), which is notably earlier than the estimate of ~22.36 Ma (95% highest posterior density interval, 16.87–28.64 Ma; Supplementary Figs. 4 and 5) based on proteomic evidence only, and more in line with some recent morphological assessments (for example, ref. [28]).

Discussion

In most respects, our higher-level results for Folivora are consistent with recently published morphology-only phylogenies, but the few ways in which they differ are critical because they have profound implications for macro-evolutionary and biogeographical inference. Harmonization of morphological and molecular datasets is complicated, as the molecular results imply that traditional clades exhibit a massive amount of unrecognized homoplasy—or equally unrecognized plesiomorphies, incorrectly interpreted as (syn)apomorphies. Molecular analyses are of course subject to the same challenges, especially in contexts like the present in which samples sizes and information content are limited. It is already widely appreciated that genomic information is exceptionally useful for testing phylogenetic hypotheses; so is proteomic information, especially when it can be shown to be highly congruent with genetic indicators of relationship[53]. Together, as illustrated here, they provide a strong basis for formulation of the following evolutionary hypotheses.

***Choloepus* is a mylodontoid.** That the two-toed sloth may be closer to traditional mylodontids than to megalonychids, a possibility occasionally raised in morphological studies[16,23,57], has been consistently found in recent aDNA investigations[33,34,36,37,39,40]. Due to the limited number of extinct taxa included in those investigations, the exact nature of their relationship has remained indeterminate. However, the multiple tests of phylogenetic relationships and broad taxonomic sampling used in the present study substantiate the conclusion that *Choloepus* is indeed a mylodontoid.

Given the recent ages of all of the taxa investigated, coupled with low rates of sequence evolution, it is unsurprising that divergence estimates based on proteins alone suggest an early/middle Miocene origin for Scelidotheriidae + Mylodontidae (including *Choloepus*). Inclusion of genomic data helps to push these estimates back to the earliest Miocene, but it should be noted that a number of mylodontoid sloths of late Oligocene to late Miocene/early Pliocene age do not fit neatly into better-defined clades. In the past, these taxa were occasionally gathered[9,20,38] into the probably non-monophyletic grouping Orophodontidae. It would be interesting to

Table 2 | Selected divergence time estimates from BEAST analyses using different combinations of taxa and data (see Results, 'Phylogenetic reconstruction', and Supplementary Information)

	Protein only		Mitochondrial DNA + protein
Clade	**Xenarthra**	**Folivora**	**Xenarthra**
Crown Xenarthra	62.0 (57.6–62.8)	–	62.6 (58.0–70.2)
Pilosa	50.4 (37.4–62.8)	–	52.1 (35.8–64.8)
Folivora	26.4 (18.0–36.0)	23.4 (14.9–33.9)	31.2 (21.1–41.4)
Megalocnidae	9.9 (3.8–17.8)	7.7 (3.4–13.0)	12.7 (4.4–22.6)
Megatherioidea + Megalocnidae	–	19.4 (12.8–27.8)	–
Megatherioidea + Mylodontoidea	22.7 (16.1–31.0)	–	26.9 (17.2–34.4)
Megatherioidea	15.7 (10.7–21.8)	13.9 (9.4–19.4)	23.0 (14.0–30.1
Megalonyx + *Bradypus*	11.1 (8.4–15.0)	10.5 (8.4–14.1)	18.4 (8.4–25.2)
Bradypus spp.	–	–	16.0 (1.5–22.1)
Megatherium + *Nothrotheriops*	12.3 (8.4 –17.7)	10.9 (7.8–15.1)	16.7 (9.9–23.6)
Mylodontoidea	15.3 (9.8–21.4)	15.4 (8.9–23.4)	22.2 (14.7–30.0)
Choloepus + Mylodontidae	12.03 (7.3–17.2)	10.5 (6.2–15.9)	20.5 (13.8–27.9)
Choloepus spp.	–	–	6.8 (2.6–11.8)

Note that, although consistently recovered as monophyletic, the position of Megalocnidae shifted among analyses, falling alternately as sister to all other Folivora (Xenarthra) or Megatherioidea (Folivora). Per-taxon estimate is cited as mean plus confidence interval (in parentheses).

know, on the basis of molecular evidence, whether the inclusion of a putative orophodontid would affect the placement of *Choloepus*, possibly moving it stemward (Supplementary Fig. 6), or help refine divergence time estimates at the base of *Mylodontoidea. At present there is no evidence on point; however, the youngest of these ambiguously placed taxa, *Octodontobradys*, is late Miocene/early Pliocene in age[58]—young enough to stand a chance of coming within the range of proteomic methods as these continue to improve.

Megalocnid sloths are monophyletic, and are not part of traditional Megalonychidae. Antillean sloths have had a complex taxonomic history[7]. In the past, this geographical grouping of folivorans was sometimes regarded as diphyletic, with different island taxa having diverged from different mainland antecedents[3,7,38,58,59]. Diphyly now seems unlikely on the basis of both our molecular clock results (Fig. 4; see also Supplementary Figs. 4 and 5) and recent morphology-based studies[16,24]. Although within-clade relationships are poorly resolved (cf. paraphyletic *Neocnus*), the Antillean clade as a whole resolves as strongly monophyletic (PP > 0.99). In light of this fact, as well as clade antiquity, it is appropriate to remove Megalocninae from traditional Megalonychidae and raise it to family level (*Megalocnidae).

***Megalonyx* and *Bradypus* are megatherioids.** Although recent morphology-oriented cladistic studies have usually recovered *Bradypus* as sister to all other folivorans[8–10,16], genomic approaches[33,34,37] have consistently paired the three-toed sloth with the extinct North American Pleistocene megatherioid *Nothrotheriops*. On this point the proteomic data presented here are fully compliant with the genomic evidence and support rejection of the inference[9,16] that Bradypodoidea (that is, *Bradypus*) is sister to traditional Megatherioidea + Mylodontoidea, as tested by both parsimony (13 additional steps) and Bayesian inference (2 × lnBayes Factor = 6.72, support *Strong*). Equally controversial is the sister group relationship detected between *Bradypus* + *Megalonyx* (PP = 0.89–0.98; Fig. 4; see Supplementary Fig. 2). Although well supported in analyses of both collagen-only and combined proteomic + genomic data, this remains a surprising finding inasmuch as such an association has never been reported in any taxon-rich phylogenetic study emphasizing morphology. While both the three-toed sloth and *Megalonyx* are likely to be megatherioids cladistically, settling their deeper relationships will require substantially more data than are currently available.

That none of the Antillean sloths used in this study showed any proteomic affinity for *Megalonyx* is also surprising, because much of what has been understood to characterize non-South American Megalonychidae on morphological grounds was based on Antillean species, the fossils of which tend to be far more complete than those of most other taxa conventionally included in this family[12,16,17]. To resolve this conflict, additional high-quality data will be required, genomic and proteomic as well as phenomic. The only certainty at present is that, if *Choloepus* is excluded, Megalonychidae must now be relegated to the list of formerly diverse but now completely extinct folivoran families.

The West Indies may have been colonized early. An early appearance of megalocnid sloths in the West Indies has been proposed on general palaeobiogeographical grounds[3,10,17,23,60], but at present the only pre-Quaternary fossil evidence for Antillean folivorans consists of a morphologically inconclusive partial femur from the early Oligocene (~31 Ma) Yauco Formation of Puerto Rico[61] and unassociated remains attributable to a definite folivoran, *Imagocnus zazae*, from the late early Miocene (~17.5 Ma) Lagunitas Formation of Cuba[5]. Although 'megalonychid' affinities have been assumed for both on biogeographical grounds, now no longer applicable, neither has been included in formal phylogenetic analyses and their placement within Folivora remains uncertain.

The presence of sloths in the West Indies at least as early as the early Miocene is congruent with our mean age estimate (31.2 Ma; Fig. 4 and Table 2) for the last common ancestor of sloths sampled in this study. This inference is also roughly consistent with the GAARlandia dispersal hypothesis[5,62], which holds that northwestern South America and the Greater Antilles were briefly in land connection during the Eocene–Oligocene transition. Without going beyond the very slim body of molecular evidence currently available, there is now at least some basis for hypothesizing that *Megalocnidae might represent an in situ Antillean radiation that was emplaced on the islands during the earliest phases of the evolution of the folivoran crown-group—much earlier than previously

thought and inconsistent with the hypothesis of a Patagonian origin for Folivora as a whole[9]. If it proves possible to acquire genomic information from Greater Antillean sloth taxa known to have survived into the mid-Holocene[63], we may expect more light to be shed on megalocnid origins.

Systematic repositioning of *Bradypus, Choloepus* and megalocnid sloths also permits a better understanding of how often 'extreme' arboreality arose during folivoran evolution. The living tree sloths are uniquely defined among extant vertebrates by a combination of relatively rigid, hook-like hands and feet, marked limb mobility, extremely long arms and powerful flexion capabilities in proximal limb joints[19]. None of the West Indian sloths possessed all of these osteological traits but, importantly, some came close—notably the Puerco Rican species *Acratocnus odontrigonus*, which may have been technically capable of hand- and foot-suspension but probably did not perform the 'upside-down' form of locomotion characteristic of extant sloths[7,14]. Remains assigned to the early Miocene Patagonian sloth *Eucholoeops*, possibly part of a clade ancestral to the Antillean radiation, also display many features consistent with highly developed arboreality[14,18]. Our phylogenetic results suggest that evolutionary experiments connected with life in the trees probably occurred multiple times, and early on, in folivore evolution. If so, it is puzzling that small-bodied sloths with highly mobile limbs and other arboreal adaptations are as yet unknown for the interval between the early Miocene (for example, *Eucholoeops*) and the Quaternary (for example, *Diabolotherium*)[18]. It is possible that their absence is only apparent; that is, highly arboreal taxa existed throughout this time period, but were confined to heavily forested tropical environments that did not favour fossilization (for example, mid-Cenozoic proto-Amazonia[64,65]).

The advent of molecular resources providing new information on both extinct and extant species offers new ways of testing hypotheses about relationships that, in the past, were by necessity based on morphological data alone. Thanks to ongoing improvements in instrumentation and applicable software, the future for palaeoproteomics should be bright if it can continue to make major contributions to solving difficult questions such as those explored here.

A new aDNA study[66] of folivoran phylogeny, published as this paper was going to press, reaches conclusions almost identical to ours regarding the evolutionary relationships of living tree sloths and the phylogenetic distinctiveness of the West Indian radiation. Because the taxonomic distribution of sampled species is not identical in the two studies, there are some minor differences in lower-level relationships and estimated divergence times. However, their detailed agreement overall supports the argument that high-quality protein sequence information is a reliable source of evidence for reconstruction of phylogenetic relationships.

Methods

Proteomic analyses. The five-digit codes following taxon names in this section refer to laboratory sample ID numbers referenced in Table 1.

AAR. Samples were prepared using a slightly modified version of the protocol in ref. [67]. A small sub-sample of bone (~1 mg) was hydrolysed in 7 M HCl (100 µl mg^{-1}) under N_2 for 18 h at 110 °C. After hydrolysis, the samples were dried overnight before rehydration in 0.01 mM L-homo-arginine as an internal standard. The samples were analysed using reversed-phase HPLC (RP–HPLC) following a slightly modified version of the protocol developed in ref. [68]. Amino acid composition and extent of racemization were used to assess samples promising for sequencing.

Sample preparation for mass spectrometry. The majority of samples (see Supplementary Table 1) were prepared using a slightly modified version of the ZooMS protocol for bone reported in ref. [43]. Bone samples (15–30 mg) were demineralized in 250 µl 0.6 M HCl for a minimum of three weeks at −20 °C. This allowed for a gentler demineralization and helped to protect any remaining collagen. After demineralization, the samples were rinsed once in 200 µl 0.01 M NaOH and three times in 200 µl 50 mM ammonium bicarbonate (Ambic). The samples were gelatinized by resuspesion in 100 µl 50 mM Ambic and heated at 65 °C for 1 h before digestion overnight at 37 °C; 50 µl of the heated sample was digested using 1 µl of 0.5 µg µl^{-1} porcine trypsin in trypsin resuspension buffer (Promega), and the other 50 µl was dried and resuspended in 50 µl 100 mM Tris solution for digestion with elastase (Worthington) at the same concentration in 10% Tris solution. Two different enzymes were used to increase the protein sequence coverage for LC–MS/MS[43,69]. Digestion was stopped by the addition of trifluoroacetic acid (TFA) at a concentration of 0.5–1% of the total solution. Peptides were desalted using zip-tips[64] and eluted in 100 µl of 50% acetonitrile (ACN)/0.1% TFA (v/v).

SDS–PAGE. Selected samples were analysed using SDS–PAGE (Table 1). This method was used on some samples, as the standard ZooMS protocol had not yielded positive results on certain samples that were deemed potentially important phylogenetically. Bone samples were crushed to ~1 µm-sized particles using a Retsch PM100 ball mill cooled with liquid nitrogen. The ball mill was cleaned with distilled water and methanol before and after each sample[70]. Nanoscale crushing allowed for the highest potential retrieval of proteomic information. Fifty milligrams of powdered sample were heated at 70 °C for 10 min in 200 µl SDS solubilizing buffer (0.5 M Tris base, 5% SDS, 130 mM DTT). Cysteines were alkylated by the addition of 6 µl 1 M indoleacetic acid at room temperature in the dark for 30 min before the addition of 200 µl of dye solution (0.05% bromophenol blue, 5% glycerol). Samples (20 µl) were run on a Bis-Tris gel (NuPAGE) for 10 min to concentrate them into a gel plug, which was briefly washed in a fixing solution (16% methanol, 10% acetic acid) before being washed twice in boiling water. The gel was stained using Coomassie stain.

The gel plug was cut with a scalpel into approximately 1 mm-sized cubes in a fume hood, and the gel cubes for each sample placed in a separate Eppendorf tube. The gel pieces were washed in a de-staining solution (66% ammonium bicarbonate, 33% ACN) until no more dye could be seen, before being washed in the following solvents for 10 min per solvent: acetonitrile (ACN), HPLC-grade water, and ACN and 50 mM ammonium bicarbonate[71]. The samples were digested overnight with 100 µl 3.125 µg µl^{-1} trypsin in 50 mM ammonium bicarbonate at 37 °C, and the tryptic digest was then pipetted into a cleaned Eppendorf tube. Next,70% ACN/1.7% formic acid/0.1% TFA (100 µl)was added to the gel pieces and the gel was heated at 37 °C for 1 h, with the supernatant being collected and added to the tryptic digest. This step was repeated sequentially with 100 mM triethyl ammonium bicarbonate and ACN. The extracted peptides were dried and then resuspended in 5% formic acid/0.1% TFA desalted and purified on C18 membranes (Empore), before elution in 80% ACN/0.5% acetic acid. The purified peptides were spun to dryness ready for LC–MS/MS analysis.

MALDI-TOF MS. A 1-µl volume of sample was spotted in triplicate onto an MTP384 Bruker ground-steel MALDI target plate. Next, 1 µl of α-cyano-4-hydroxycinnamic acid matrix solution (1% in 50% ACN/0.1% TFA (v/v/v)) was added to each sample spot and mixed with the sample[43]. All samples were analysed on a Bruker Ultraflex MALDI-TOF mass spectrometer in triplicate.

LC–MS/MS. Most samples were analysed at the Discovery Proteomic Facility (DPF) at Oxford, UK (Table 1). *Choloepus* ID 17009 and *Mylodon* ID 16222 were analysed at the Novo Nordisk Foundation Centre for Protein Research (NNFCPR), University of Copenhagen. The *Megalonyx* sample (ID 16849) was run at the Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, Rockefeller University.

At DPF, sample batches were analysed on an Orbitrap Fusion Lumos or Q-Exactive with identical front-end separation, employing an EASY-Spray column (ES803, 500 mm × 75 µm, Thermo) and a gradient of 2–35% ACN in 0.1% formic acid/5% DMSO over 60 min. On the Fusion Lumos, MS1 resolution was set to 120,000 with an automatic gain control (AGC) target of 400,000. MS2 spectra were acquired in TopSpeed mode (3-s duty cycle) in the linear ion trap (rapid scan mode) for up to 250 ms, with an AGC target of 4,000 and fragmentation in collision-induced dissociation mode (35% normalized collision energy). MS1 resolution on the Q-Exactive was set to 70,000 with an AGC target of 3×10^6. MS2 spectra for up to 15 precursors were acquired with a resolution of 17,500 and an AGC target of 1×10^5 for up to 128 ms and 28% normalized collision energy (higher-energy collision dissociation). On both instruments, precursors were excluded for 27 s from re-selection.

At NNFCPR, dried peptides were resuspended in 50 µl of 80% ACN and 0.1% formic acid before transfer to a 96-well plate and placed in a vacuum centrifuge at 40 °C until approximately 3 µl of solution remained. The samples were rehydrated with either 5 or 10 µl (*Mylodon* ID 16222 and *Choloepus* ID 17009, respectively) of 0.1% TFA and 5% ACN. Samples were separated on a 15-cm column (75 µm inner diameter) in-house laser pulled and packed with 1.9 µm C18 beads (Dr. Maisch) on an EASY-nLC 1000 (Proxeon) connected to a Q-Exactive HF (Thermo Scientific) on a 77-min gradient; 5 µl of sample was injected. Buffer A was milliQ water. The peptides were separated with increasing amounts of buffer B (80% ACN and 0.1% formic acid), rising from 5 to 80% over an 80-min gradient, and a flow rate of 250 nl min^{-1}. In addition, a wash-blank injecting 2 µl 0.1% TFA and 5% ACN was run between each sample to hinder cross-contamination.

The Q-Exactive HF was operated in data-dependent top-10 mode. Full-scan mass spectra (350–1400 m/z) were recorded at a resolution of 120,000 at 200 m/z

with a target value of 3×10^6 and a maximum injection time of 25 ms for *Choloepus* ID 17009 and 45 ms for *Mylodon* ID 16222. Fragment ions were recorded with a maximum ion injection time set to 108 ms and a target value set to 2×10^5, and recorded at a resolution of 60,000 for *Choloepus* ID 17009 and 30,000 for *Mylodon* ID 16222. Normalized collision energy was set at 28% and the isolation window was 1.2 *m/z*, with the dynamic exclusion set to 20 s.

At Rockefeller University, peptides were resuspended in 20 µl 5% methanol and 0.2% formic acid. Next, 10 µl were loaded onto an EASY-Spray column (Thermo Fisher Scientific ES800, 15 cm × 75 µm ID, PepMap C18, 3 µm) with an EASY-nLC 1200 and separated over a 120-min gradient of 2–32% Solvent B (Solvent A, 0.1% formic acid in water, Solvent B, 0.1% formic acid, 95% acetonitrile) during online electrospray ionization–MS and MS/MS analyses with a Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific). MS/MS analyses of the top 25 precursors in each full scan (300–1700 *m/z*) used the following parameters: resolution, 17,500 (at 200 Th); AGC target, 2×10^5; maximum injection time, 200 ms; isolation width, 2.0 *m/z*; normalized collision energy, 24%.

Protein sequence analysis. The LC–MS/MS raw files were converted to Mascot generic files using Proteowizard[72] and searched against a mammal collagen database that included common contaminants (http://www.thegpm.org/crap/) in PEAKS v.7.5. Mass tolerances were set at 0.5 Da for the fragment ions and 10 ppm for precursor ions, and up to three missed cleavages were permitted. Searches allowed various post-translational modifications (PTMs) including oxidation of the amino acids methionine, histidine, and tryptophan (+15.99), hydroxylation of proline (also +15.99), deamidation of asparagine and glutamine (+0.98), pyro-glu from glutamic acid (−18.01) and carbamidomethylation (+57.02) which occurs as part of the sample preparation. A maximum of three PTMs was allowed per peptide. Protein tolerances were set at 0.5% false discovery rate, >50% average local confidence (de novo only) and $-\log_{10}P$ score ≥ 20.

Sequences of both *COL 1A1* and *COL 1A2* were concatenated using previously published mammal collagen consensus sequences derived from NCBI, including sequences for the xenarthrans *Dasypus novemcinctus* (nine-banded armadillo; GenBank, No. XP_004470764), *Cyclopes didactylus* (silky anteater; Uniprot, No. COHJP1/COHJP2) and *Lestodon armatus* (extinct mylodontoid sloth, ref. [44]). Telopeptides very rarely survive in fossil samples and so these were removed from all sequences. The amino acids isoleucine and leucine cannot be differentiated using low-energy tandem mass spectrometry and de novo sequencing, as both are isobaric. Therefore, the identification of leucine/isoleucine was consistent throughout the sequence analyses concatenated in this study. Our approach is in line with previous phylogenetic studies using collagen as probe[43], under the assumption that MS/MS sequence variation is not interpreted as a major phylogenetic change (see Phylogenetic analyses, below).

Once a potential collagen sequence was compiled for a given sloth taxon, the sequence was added to the collagen database and the sample was rerun through PEAKS to check for coverage and sequence substitutions. Any differences noted either in the consensus sequences or between different species of sloths were inspected manually. For a difference to be considered authentic, it had to occur in more than one product ion spectrum and be covered by both b and y ions. For additional discussion, see Supplementary Table 4.

Phylogenetic analyses. Sequences developed from the MS/MS analyses were aligned in Geneious v.9.1.7 (ref. [73]) using the MUSCLE algorithm[74] with default settings, and then checked visually. Mitochondrial sequence data for extant folivorans and *M. darwinii* were obtained from ref. [34] and supplemented with protein coding sequences for *N. shastensis* from ref. [33]. Because the order of genes differs between these two alignments, we extracted and aligned genes for *Nothrotheriops* individually using MUSCLE in Geneious, checking each visually to ensure accuracy. Of the 2,096 amino acids in our alignment of the type 1 collagen molecule, 134 (6.4%) were variable and 76 (56% of variable sites, 3.6% of total) were parsimony informative for the taxa represented.

We conducted three sets of phylogenetic analyses on the resulting protein alignment (see Results, 'Phylogenetic reconstruction'). We first performed a strict parsimony analysis using PAUP v.4.0a (build 157)[75]. We employed a branch-and-bound search, with all sites treated as unordered and equally weighted. To assess clade support, we performed 10,000 bootstrap replicates using full heuristic tree searches and generated a weighted 50% majority rule consensus tree from the resulting sample of most-parsimonious bootstrapped trees.

We performed two forms of model-based phylogenetic analyses, both in a Bayesian framework. We used PartitionFinder v.2.1.1 (refs. [76,77]) to determine the most appropriate model(s) of amino acid substitution and partitioning scheme for our concatenated alignment, resulting in selection of separate Dayhoff models[78] with gamma-distributed rates for *COL 1A1* and *COL 1A2*. The first set of Bayesian phylogenetic analyses used MrBayes v.3.2.5 (ref. [79]). We performed two Markov chain Monte Carlo (MCMC) runs, each of four chains (one cold, three heated), for 10,000,000 generations, sampling from the chain every 5,000 generations. After checking for convergence of the two chains based on Gelman–Rubin statistics and ensuring that effective sample sizes for all parameters were sufficient (>200), we discarded the first 50% of each chain as burn-in, combined the remaining posterior samples and summarized these as a 50% majority rule consensus tree, with clade frequencies interpreted as posterior probabilities for a given clade. To determine whether our unconstrained topology provided a better explanation of the data than a previously proposed morphological topology[16] in which *Bradypus* is the sister lineage to all other folivorans and *Choloepus*, Megalocnidae and *Megalonyx* form a monophyletic Megalonychidae (including other taxa not referenced here), we estimated the marginal likelihood of the data on unconstrained and constrained topologies using the stepping-stone algorithm in MrBayes. We performed two runs, each with four chains (three heated, one cold) for 10,000,000 generations over 50 steps, with default settings for the alpha parameter of the beta distribution (0.4) and burn-in (−1). We calculated $2\times\mathrm{Ln}(\mathrm{Likelihood}_{\mathrm{unconstrained}}-\mathrm{Likelihood}_{\mathrm{constrained}})$ from the resulting estimates and assessed support using the scale given in ref. [80].

The fact that we cannot differentiate between isoleucine and leucine using low-energy tandem mass spectrometry creates a unique problem for model-based phylogenetic inference procedures. The standard approach in ancient protein studies[43] is to designate all sites with a molecular mass of 131.17 g mol^{-1} as leucine, but this has the potential to bias estimates of the instantaneous rate matrix, branch lengths and, possibly, topology by entirely excluding one amino acid. We investigated this by replacing all peptides coded as leucine with ambiguous codings {IL} and repeating Bayesian estimation of topology and branch lengths using MrBayes. The resulting 50% majority rule consensus tree was identical across coding schemes, and comparison of branch length estimates among analyses showed no significant deviation from 1:1 ($\mathrm{branch\ length}_{\mathrm{Leucine}}=-0.00009+\mathrm{branch\ length}_{\mathrm{ambiguous}}\times0.96$, $R^2=0.995$, $P\ll0.001$), indicating that the use of leucine is appropriate. We repeated Bayesian analyses of the combined proteomic + genomic dataset using the same settings, but with partitioning schemes and substitution models for genetic data following ref. [34].

We attempted to integrate our combined molecular dataset with a large, recently published morphological dataset (ref. [9]). The resulting majority rule consensus tree (Supplementary Fig. 6) is congruent in some respects with our molecular topologies (for example, *Choloepus* was recovered as a mylodontoid and *Bradypus* as a megatherioid), but other results repeatedly found in molecular analyses were not obtained. In particular, we recovered a strong (PP = 1.0) traditional Megalonychidae nested within Megatherioidea that included Antillean sloths minus *Choloepus*. Although the Antillean species were represented in the total dataset by proteomic sequences, genomic data were unavailable. This result suggests that the large number of morphological characters, some known to be highly homoplastic[54], were able to swamp the signal arising from the smaller proteomic dataset. While combined analysis of morphological and molecular data will ultimately be necessary to fully resolve folivoran phylogeny, this exercise suggests that it is premature to consider such simultaneous analyses reliable at this point in time.

Our MrBayes analyses sample tree topologies with branch lengths in units of substitutions per site, and thus ignore temporal information inherent in phylogenetic analysis of non-contemporaneous tips or external information about relative branch lengths that can be provided by the fossil record. We therefore also performed a series of Bayesian tree searches assuming a molecular clock under the fossilized birth–death framework[81–83], as implemented in BEAST v.2.5.1 (ref. [84]). Briefly, this framework allowed us to sample from the posterior distribution of time-scaled trees for taxa in our proteomic dataset, inferred using their sequences and stratigraphic ages, while using phylogenetically constrained fossil taxa that lack amino acid data able to provide additional information on relative branch lengths and divergence times. Our choice of fossil taxa and topological constraints broadly followed the approach undertaken in ref. [33] for sloth mitogenomes. However, our proteomic topologies raise questions about the phylogenetic positioning of some fossil folivorans that have previously been considered on morphological grounds as early representatives of Pleistocene and Holocene families. For example, some extinct folivorans, such as the Huayquerian nothrotheriid *Mionothropus*[27], can plausibly be assigned to a specific terminal branch in our proteomic topology. Others, however, are customarily assigned to clades that we failed to recover. This applies to the Santacrucian taxon *Eucholoeops*, usually interpreted as a basal megalonychid[23,85,86] and therefore as a member of a clade not found to be monophyletic in our analyses. Such issues inevitably affect efforts to calibrate the proteomic + genomic data clock and to infer divergence times. Acknowledging this, we employed a minimal set of constraints (see Supplementary Fig. 3) on the positioning of fossil folivorans in our Bayesian estimation of topology and divergence times, integrating over all possible placements of phylogenetically uncertain fossils using stratigraphic context alone when necessary. We performed analyses with and without a monophyly constrain on *Bradypus*, and results did not differ at unaffected nodes.

The use of a Bayesian approach requires the specification of prior probabilities on model parameters. We used default priors on substitution model parameters but specified the following: net diversification, ~Exp(1), yielding a broad, vague prior; turnover, ~beta(2,1), yielding high prior weight on extinction ≅ speciation; sampling probability, ~beta(2,2), yielding a humped distribution that placed most prior weight on sampling probabilities of 0.5; origin, ~U(61.5, 150), yielding a flat prior on ages older than 61.5–150 Ma. In addition, the analysis was conditioned on the number of extant taxa sampled ($\rho=0.129$ in the xenarthran proteomic analyses, $\rho=0.333$ in the folivoran proteomic analyses, $\rho=0.266$ in the combined analyses). Based on comparisons of marginal likelihoods computed

via path sampling (see Supplementary information and Supplementary Table 3), we employed a relaxed uncorrelated clock with log-normally distributed rates for proteomic and combined analyses, with an exponential prior (mean, 0.1) placed on the mean of log-normal distribution and the default gamma Γ(0.5396, 0.3819) on the standard deviation. Two MCMC analyses were run for 10 million generations each, sampling every 1,000 generations, after which fossils without data were pruned from the trees, the first 20% of the retained samples were discarded as burn-in, the samples were combined and maximum clade credibility trees constructed using the tree annotator software accompanying the BEAST suite. Runs from the prior using a fixed topology (the maximum clade credibility tree based on the pre-pruning sample) were used to confirm that divergence time estimates were not simply returning the prior.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Mass spectrometry proteomics data have been deposited with the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD012859. Collagen sequences are available on the Uniprot website (https://www.uniprot.org/); the complete list can be found in Supplementary Table 5. Phylogenetic datasets have been deposited at DataDryad (https://doi.org/10.5061/dryad.7dd64gs).

Received: 25 September 2018; Accepted: 28 April 2019;
Published online: 06 June 2019

References

1. Gardner, A. L. in *Mammals of South America* Vol. 1 (ed. Gardner, A. L.) 157–176 (Univ. of Chicago Press, 2007).
2. Nowak, R. *Walker's Mammals of the World: Monotremes, Marsupials, Afrotherians, Xenarthrans, and Sundatherians* (Johns Hopkins, 2018).
3. Kraglievich, L. Descripción de dos cráneos y otros restos del género "Pliomorphus" Ameghino, procedentes de la formación entrerriana de las barrancas del río Paraná. *Anal. Mus. Nac. Hist. Nat. Buenos Aires* **33**, 1–56 (1923).
4. Hoffstetter, R. in *Traité de Paléontologie* Vol. 6.2 (ed. Piveteau, J.) 535–636 (Masson, 1958).
5. MacPhee, R. D. E. & Iturralde-Vinent, M. A. Origin of the greater antillean land mammal fauna 1: new Tertiary land mammals from Cuba and Puerto Rico. *Am. Mus. Novit.* **3141**, 1–31 (1995).
6. Iturralde-Vinent, M. A. & MacPhee, R. D. E. Paleogeography of the Caribbean region: implications for Cenozoic biogeography. *Bull. Am. Mus. Nat. Hist.* **238**, 1–95 (1999).
7. White, J. & MacPhee, R. D. E. in *Biogeography of the West Indies: Patterns and Perspectives* 2nd edn (eds Woods, C. A. & Sergile, F. E.) 201–236 (CRC Press, 2001).
8. Gaudin T. J. & McDonald, H. G. in *The Biology of Xenarthra* (eds Vizcaíno, S. F. & Loughry, W. J.) 24–36 (Univ. Press of Florida, 2008).
9. Varela, L., Tambusso, P. S., McDonald, H. G. & Fariña, R. A. Phylogeny, macroevolutionary trends and historical biogeography of sloths: insights from a Bayesian morphological clock analysis. *Syst. Biol.* **68**, 204–218 (2018).
10. Pujos, F., De Iuliis, G. & Cartelle, C. A paleogeographic overview of tropical forest sloths: towards an understanding of the origin of extant suspensory sloths? *J. Mammal. Evol.* **24**, 19–38 (2017).
11. Patterson, B. & Pascual, R. Evolution of mammals on southern continents. *Q. Rev. Biol.* **43**, 409–451 (1968).
12. Engelmann, G. F. in *The Evolution and Ecology of Armadillos, Sloths, and Vermilinguas* (ed. Montgomery, G. G.) 195–203 (Smithsonian Institution, 1985).
13. Webb, S. D. in *The Evolution and Ecology of Armadillos, Sloths, and Vermilinguas* (ed. Montgomery, G. G.) 105–112 (Smithsonian Institution Press, 1985).
14. White, J. Indicators of locomotor habits in xenarthrans: evidence of locomotor heterogeneity among fossil sloths. *J. Vertebr. Paleontol.* **13**, 230–242 (1993).
15. Delsuc, F., Catzeflis, F. M., Stanhope, M. J. & Douzery, E. J. P. The evolution of armadillos, anteaters and sloths depicted by nuclear and mitochondrial phylogenies: implications for the status of the enigmatic fossil Eurotamandua. *Proc. R. Soc. B* **268**, 1605–1615 (2001).
16. Gaudin, T. J. Phylogenetic relationships among sloths (Mammalia, Xenarthra, Tardigrada): the craniodental evidence. *Zool. J. Linn. Soc.* **140**, 255–305 (2004).
17. McDonald, H. G. & De Iuliis, G. in *The Biology of Xenarthra* (eds Vizcaino, S. F. & Loughry, W. J.) 39–55 (Univ. Press of Florida, 2008).
18. Pujos, F., Gaudin, T. J., De Iuliis, G. & Cartelle, C. Recent advances on variability, morpho-functional adaptations, dental terminology, and evolution of sloths. *J. Mamm. Evol.* **19**, 159–169 (2012).
19. Nyakatura, J. A. The convergent evolution of suspensory posture and locomotion in tree sloths. *J. Mamm. Evol.* **19**, 225–234 (2012).
20. Patterson, B., Turnbull, W. D., Segall, W. & Gaudin, T. J. The ear region in xenarthrans (= Edentata: Mammalia). Part II. Pilosa (sloths, anteaters), palaeanodonts, and a miscellany. *Fieldiana Geol.* **24**, 1–79 (1992).
21. Pujos, F. *Megatherium celendinense* sp. nov. from the Pleistocene of Peruvian Andes and the Megatheriine phylogenetic relationship. *Palaeontology* **49**, 285–306 (2006).
22. Pujos, F., De Iuliis, G. & Mamani Quispe, B. *Hiskatherium saintandrei*, gen. et sp. nov.: an unusual sloth from the Santacrucian of Quebrada Honda (Bolivia) and an overview of middle Miocene, small megatherioids. *J. Vert. Paleontol.* **31**, 1131–1149 (2011).
23. McDonald, H. G., Rincón, A. D. & Gaudin, T. J. A new genus of megalonychid sloth (Mammalia, Xenarthra) from the late Pleistocene (lujanian) of Sierra de Perija, Zulia State, Venezuela. *J. Vert. Paleontol.* **33**, 1226–1238 (2013).
24. McDonald, H. G. & Carranza-Castaneda, O. Increased xenarthran diversity of the great American biotic interchange: a new genus and species of ground sloth (Mammalia, Xenarthra, Megalonychidae) from the Hemphillian (late Miocene) of Jalisco, Mexico. *J. Paleontol.* **91**, 1–14 (2017).
25. Brandoni, D. A new genus of Megalonychidae (Mammalia, Xenarthra) from the late Miocene of Argentina. *Rev. Bras. Paleontol.* **17**, 33–42 (2014).
26. Brandoni, D. The Megalonychidae (Xenarthra, Tardigrada) from the late Miocene of Entre Ríos Province, Argentina, with remarks on their systematics and biogeography. *Geobios* **44**, 33–44 (2011).
27. De Iuliis, G., Gaudin, T. J. & Vicars, M. J. A new genus and species of nothrotheriid sloth (Xenarthra, Tardigrada, Nothrotheriidae) from the late Miocene (Huayquerian) of Peru. *Palaeontology* **54**, 171–205 (2011).
28. Gaudin, T. J. & Croft, D. Paleogene Xenarthra and the evolution of South American mammals. *J. Mamm.* **96**, 622–634 (2015).
29. Rincón, A. D., Solórzano, A., McDonald, H. G. & Montellano-Ballesteros, M. Two new megalonychid sloths (Mammalia: Xenarthra) from the Urumaco Formation (late Miocene), and their phylogenetic affinities. *J. Syst. Palaeontol.* **17**, 409–421 (2019).
30. Boscaini, A., Gaudin, T. J., Mamani Quispe, B., Antoine, P.-O. & Pujos, F. New well-preserved craniodental remains of *Simomylodon uccasamamensis* (Xenarthra, Mylodontidae) from the Pliocene of the Bolivian Altiplano: phylogenetic, chronostratigraphic and paleobiogeographic implications. *Zool. J. Linn. Soc.* **185**, 459–486 (2019).
31. McDonald, H. G. & De Iuliis, G. in *The Biology of Xenarthra* (eds Vizcaino, S. F. & Loughry, W. J.) 39–55 (Univ. Press of Florida, 2008).
32. Delsuc, F. & Douzery, E. J. P. in *The Biology of Xenarthra* (eds Vizcaino, S. F. & Loughry, W. J.) 11–23 (Univ. Press of Florida, 2008).
33. Slater, G. et al. Evolutionary relationships among extinct and extant sloths: the evidence of mitogenomes and retroviruses. *Genome Biol. Evol.* **8**, 607–621 (2016).
34. Delsuc, F. et al. Resolving the phylogenetic position of Darwin's extinct ground sloth (*Mylodon darwinii*) using mitogenomic and nuclear exon data. *Proc. R. Soc. B* **285**, 20180214 (2018).
35. Moraes-Barros, N., Silva, J. A. & Morgante, J. S. Morphology, molecular phylogeny, and taxonomic inconsistencies in the study of *Bradypus* sloths (Pilosa: Bradypodidae). *J. Mammal.* **92**, 86–100 (2011).
36. Poinar, H. N. et al. Molecular coproscopy: dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. *Science* **281**, 402–406 (1998).
37. Greenwood, A. D., Castresana, J., Feldmaier-Fuchs, G. & Pääbo, S. A molecular phylogeny of two extinct sloths. *Mol. Phylogenet. Evol.* **18**, 94–103 (2001).
38. McKenna, M. C. & Bell, S. K. *Classification of Mammals above the Species Level* (Columbia Univ. Press, 1997).
39. Höss, M., Dilling, A., Currant, A. & Pääbo, S. Molecular phylogeny of the extinct ground sloth *Mylodon darwinii*. *Proc. Natl Acad. Sci. USA* **93**, 181–185 (1996).
40. Hofreiter, M., Betancourt, J. L., Sbriller, A. P., Markgraf, V. & McDonald, H. G. Phylogeny, diet, and habitat of an extinct ground sloth from Cuchillo Cura, Neuquen Province, southwest Argentina. *Quat. Res.* **59**, 364–378 (2003).
41. Welker, F. et al. Middle Pleistocene protein sequences from the rhinoceros genus *Stephanorhinus* and the phylogeny of extant and extinct middle/late Pleistocene Rhinocerotidae. *PeerJ* **5**, e3033 (2017).
42. Welker, F. et al. Palaeoproteomic evidence identifies archaic hominins associated with the Châtelperronian at the Grotte du Renne. *Proc. Natl Acad. Sci. USA* **113**, 11162–11167 (2016).
43. Welker, F. et al. Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. *Nature* **522**, 81–84 (2015).
44. Buckley, M. et al. Collagen sequence analysis of the extinct giant ground sloths *Lestodon* and *Megatherium*. *PloS ONE* **10**, e0144793 (2015).
45. Dobberstein, R. C. et al. Archaeological collagen: why worry about collagen diagenesis? *Archaeol. Anthropol. Sci.* **1**, 31–42 (2009).
46. Buckley, M. & Collins, M. J. Collagen survival and its use for species identification in Holocene-Lower Pleistocene bone fragments from British archaeological and palaeontological sites. *Antiqua* **1**, e1 (2011).
47. Buckley, M. & Wadsworth, C. Proteome degradation in ancient bone: diagenesis and phylogenetic potential. *Palaeogeog. Palaeoclimatol. Palaeoecol.* **416**, 69–79 (2014).

48. Rybczynski, N. et al. Mid-Pliocene warm-period deposits in the High Arctic yield insight into camel evolution. *Nat. Comm.* **4**, 1550 (2013).
49. Allentoft, M. E. et al. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc. R. Soc. B* **279**, 4724–4733 (2012).
50. Orlando, L. et al. Recalibrating *Equus* evolution using the genome sequence of an early middle Pleistocene horse. *Nature* **499**, 74–78 (2013).
51. Presslee, S. et al. Radiocarbon dating and proteomic analysis of highly purified bone collagen derived from Rancho la Brea mammal fossils. *Society of Vertebrate Paleontology Annual Meeting Program* 208 (2016).
52. Tuross, N. & Stathoplos, L. in *Methods in Enzymology* Vol. 224 (eds Zimmer A., White, T. J., Cann, R. L. & Wilson, A. C.) 121–129 (Academic Press, 1993).
53. Westbury, M. et al. A mitogenomic timetree for Darwin's enigmatic "transitional" South American mammal, *Macrauchenia Patachonica*. *Nat. Commun.* **8**, 15951 (2017).
54. Hautier, L., Gomes Rodrigues, H., Billet, G. & Asher, R. J. The hidden teeth of sloths: evolutionary vestiges and the development of a simplified dentition. *Sci. Rep.* **6**, 27763 (2016).
55. Cione, A. L. & Tonni, E. P. in *Quaternary of South America Antarctic Península* (eds Tonni, E. P. & Cione, A. L.) 23–51 (Balkema,1999).
56. Cartelle, C., De Iuliis, G. & Ferreira, R. L. Systematic revision of tropical Brazilian scelidotheriines sloths (Xenarthra, Mylodontoidea). *J. Vertebr. Paleontol.* **29**, 555–566 (2009).
57. Guth, C. *La Région Temporale des Edentés* (Imprimerie Jeanne d'Arc Le Puy, 1961).
58. Guilherme, E., Bocquentin, J. & Porto, A. S. A new specimen of the genus *Octodontobradys* (Orophodontidae, Octodontobradyinae) from the late Miocene-Pliocene of the southwestern Amazon Basin, Brazil. *Anu. ár. Inst. Geociências* **34**, 64–71 (2011).
59. Varona, L. *Catálogo de los Mamíferos Vivientes y Extinguidos de las Antillas* (Academia de Ciencias de Cuba, 1974).
60. Webb, S. D. & Perrigo, S. in *The Evolution and Ecology of Armadillos, Sloths, and Vermilinguas* (ed. Montgomery, G. G.) 113–120 (Smithsonian Institution Press, 1985).
61. MacPhee, R. D. E., Iturralde-Vinent, M. A. & Gaffney, E. S. Domo de Zaza: an early Miocene vertebrate locality in south-central Cuba, with notes on the tectonic evolution of Puerto Rico and Mona Passage. *Am. Mus. Novit.* **3394**, 1–42 (2003).
62. Tong, Y. F. et al. Huntsmen of the Caribbean: multiple tests of the GAARlandia hypothesis. *Mol. Phylogenet. Evol.* **130**, 259–268 (2019).
63. Steadman, D. W. et al. Asynchronous extinction of late Quaternary sloths on continents and islands. *Proc. Natl Acad. Sci. USA* **102**, 11763–11768 (2005).
64. Hoorn, C. et al. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330**, 927–931 (2010).
65. Tejada-Lara, J. V. et al. Life in proto-Amazonia: Middle Miocene mammals from the Fitzcarrald Arch (Peruvian Amazonia). *Palaeontology* **58**, 341–378 (2015).
66. Delsuc, F. et al. Ancient mitogenomics rewrites the evolutionary history and biogeography of sloths. *Curr. Biol.* https://doi.org/10.1016/j.cub.2019.05.043 (2019).
67. Penkman, K., Kaufman, D. S., Maddy, D. & Collins, M. J. Closed-system behaviour of the intra-crystalline fraction of amino acids in mollusc shells. *Quat. Geochronol.* **3**, 2–25 (2008).
68. Kaufman, D. S. & Manley, W. F. A new procedure for determining DL amino acid ratios in fossils using reverse phase liquid chromatography. *Quat. Sci. Rev.* **17**, 987–1000 (1998).
69. Demarchi, B. et al. Protein sequences bound to mineral surfaces persist into deep time. *eLife* **5**, e17092 (2016).
70. Kontopoulos, I., Presslee, S., Penkman, K. & Collins, M. J. Preparation of bone powder for FTIR-ATR analysis: the particle size effect. *Vib. Spectrosc.* **99**, 167–177 (2018).
71. Van Doorn, N. L., Hollund, H. & Collins, M. J. A novel and non-destructive approach for ZooMS analysis: ammonium bicarbonate buffer extraction. *Archaeol. Anthropol. Sci.* **3**, 281–289 (2011).
72. Chambers, M. C. et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30**, 918–920 (2012).
73. Kearse, M. et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
74. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
75. Swofford, D. L. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods) Version 4 (Sinauer Associates, 2002).
76. Lanfear, R., Calcott, B., Ho, S. Y. W. & Guindon, S. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* **29**, 1695–1701 (2012).
77. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **34**, 772–773 (2017).
78. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. in *Atlas of Protein Sequence and Structure* Vol. 5 (ed. Dayhoff, M. O.) 345–352 (National Biomedical Research Foundation, 1978).
79. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
80. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
81. Heath, T. A., Huelsenbeck, J. P. & Stadler, T. The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proc. Natl Acad. Sci. USA.* **111**, 2957–2966 (2014).
82. Gavryushkina, A., Welch, D., Stadler, T. & Drummond, A. J. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Comput. Biol.* **10**, e1003919 (2014).
83. Gavryushkina, A. et al. Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Syst. Biol.* **66**, 57–73 (2017).
84. Bouckaert, R. et al. BEAST 2: a software platform for bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
85. De Iuliis, G., Pujos, F., Toledo, N., Bargo, M. S. & Vizcaíno, S. F. *Eucholoeops ameghino*, 1887 (Xenarthra, Tardigrada, Megalonychidae) from the Santa Cruz Formation, Argentine Patagonia: implications for the systematics of santacrucian sloths. *Geodiversitas* **36**, 209–255 (2014).
86. Hirschfeld, S. E. & Webb, S. D. Plio-Pleistocene megalonychid sloths of North America. *Bull. Fla. Mus. Nat. Hist.* **12**, 213–294 (1968).

Acknowledgements

We thank the curatorial staffs of the following museums and private collections for permission to sample specimens in their care: AMNH-M, American Museum of Natural History (Mammalogy), New York, USA; AMNH-P, American Museum of Natural History (Paleontology), New York, USA; CIV, Iota Quatro faunal collection, courtesy of Lazaro Vinola; El Trebol faunal collection, Bariloche, Argentina; FR, Forest Reserve (Trinidad) faunal collection currently housed in Department of Mammalogy, AMNH, New York, USA; IANIGLA-PV, Instituto Argentino de Nivología, Glaciología y Ciencias Ambientales, CCT-CONICET-Mendoza, Mendoza, Argentina; MACN-PV, Museo Argentino de Ciencias Naturales 'Bernardino Rivadavia' (Sección Paleovertebrados), Buenos Aires, Argentina; MAPBAR, Museo de la Asociación Paleontológica Bariloche (APB), prov. Río Negro, Argentina; MMP, Museo Municipal de Ciencias Naturales 'Lorenzo Scaglia' Mar del Plata, prov. Buenos Aires, Argentina; MNHN SAO, Muséum national d'Histoire naturelle, Paris, France; MPS, Museo Paleontológico 'Fray Manuel de Torres', San Pedro, prov. Buenos Aires, Argentina; MUSM, Museo de Historia Natural de la Universidad Nacional Mayor de San Marcos, Lima, Peru; NYSM VP, New York State Museum (Vertebrate Paleontology), Albany, USA; RM, Cuban faunal collection currently housed in Department of Mammalogy, AMNH, New York, USA; UF, University of Florida, Natural History Museum of Florida (Vertebrate Paleontology), Gainesville, USA; UMAG ah, Instituto de La Patagonia, Universidad de Magallanes, Punta Arenas, Chile; USNM, United States National Museum of Natural History (Paleobiology), Washington DC, USA. Samples of specimens housed in Argentinian collections were sampled before 2009. S.P. would like to thank B. Demarchi for useful discussion and support. The authors thank the National Science Foundation for grants (No. OPP 0636639 to R.D.E.M. and No. DEB 1547414 to R.D.E.M., M.C. and K.P.).

Author contributions

R.D.E.M., M.C. and S.P. conceived the project. S.P. undertook AAR, proteomic analysis and concatenated collagen sequences, with laboratory and technical assistance from R.F., J.V.O., K.M., M.M., M.C., K.P. and B.T.C. G.J.S. conducted phylogenetic analyses. F.P. and A.M.F. supplied palaeontological information. A.K., M.T., F.S., M.L., A.H., R.F., J.B., J.L.L., F.M.M., R.S.G., M.R., A.G., C.d.M. and J.S. supplied fossil samples, locality information, dating, species identifications and commentary on the manuscript. R.D.E.M., S.P. and G.J.S. wrote the manuscript, with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-019-0909-z.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to R.D.E.M.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): ROSS D. E. MACPHEE

Last updated by author(s): Apr 8, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	None used
Data analysis	Proteomics: SDS PAGE, Proteowizard, PEAKS v. 7.5 Phylogenetics: Geneious v. 9.1.7, PAUP v. 4.0a (build 157), PartitionFinder v. 2.1.1, MrBayes v 3.2.5, BEAST v2.5.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD012859. Collagen sequences are available on the Uniprot website (https://www.uniprot.org/) and SI Table S5. Phylogenetic datasets have been deposited on DataDryad (doi:10.5061/dryad.7dd64gs). There are no post-publication restrictions on data availability.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Proteomic/phylogenetic study of extinct and extant sloths using sequence data developed from analysis of bone collagen with LC MS/MS. Effort directed toward testing the placement of the extant tree sloths with their sister groups using combinations of proteomic and published genomic and phenomic evidence
Research sample	To undertake this study it was necessary to utilise a large number of fossil samples because experience shows that only a proportion will provide high-quality sequence data. We sampled 120 elements in various museums and institutes, and 34 (28%) performed well enough for inclusion in the study.
Sampling strategy	There is no method of determining the likelihood that a bone will produce high quality collagen without sampling it first. However, once that step was undertake, we ascertained likely quality with an utilized amino acid racemization test
Data collection	Protein sequence information was collected by LC MS/MS using MALDI-ToF (Matrix-Assisted Laser Desorption/Ionization Time-of-Flight) mass spectrometry
Timing and spatial scale	N/A
Data exclusions	Samples returning poor AAR profiles were not used. Of those with acceptable profiles, the best sample per taxon was selected for LC MS/MS (15 in all).
Reproducibility	This topic is discussed in SI under the heading "Utilization of database searching vs. de novo sequencing"
Randomization	Not relevant because analysis could only proceed on samples that tested positively for collagen
Blinding	Not relevant because analysis could only proceed on samples that tested positively for collagen

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☐	☒ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Palaeontology

Specimen provenance	No specimens were collected for this paper (sampling only). All samples are referenced in SI Table S1, with museums/institutions of origin.
Specimen deposition	All samples were consumed during the course of analyses
Dating methods	SI Table S2 provides all information supplied by lab undertaking the radiocarbon dating of samples (Keck-CCAMS Group, Earth System Science Department, University of California, Irvine)

☒ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.